FIG. 7 gives a more detailed description of a preferred implementation of step 60 of FIG. 6, subject to the constraint of symmetry, just given. Where $2q$ communities are to be identified, the method of the invention may conveniently be implemented to follow the symmetry.

The definition of the 2q communities proceeds as a series of q iterations of a sequence of steps, where each iteration produces two communities, one for pages with large coordinates and one for pages with small coordinates. Here, "smallest" is taken to mean most negative, i.e., having the largest negative magnitude.

The iteration is expressed, in step 62, as a FOR loop to be executed q number of times.

In step 64, a community, indexed as community $2i$-1 (for $1<i<q$), is defined, by choosing k pages with largest coordinates in the vector H[i] as hubs (step 66), and choosing k pages with largest coordinates in the vector H[i] as hubs (step 66).

Next, in step 70, a community, indexed as community $2i$ (for $1<i<q$), is defined, by choosing k pages with smallest coordinates in the vector H[i] as hubs (step 66), and choosing k pages with smallest coordinates in the vector H[i] as hubs (step 66).

This completes an iteration. In successive iterations, the pages with progressively smaller coordinates are used to define the hubs and authorities for odd-indexed communities, and the pages with progressively larger coordinates are used to define the hubs and authorities for even-indexed communities, until all $2q$ of the communities have been defined.

MATHEMATICAL INTERPRETATION

The discussion which follows will present a somewhat more theoretical treatment of the concepts relating to hubs and authorities, which have been discussed above. Certain aspects of the discussion may be foreseen from the existing literature.

The hub and authority vectors H and A correspond to the principal eigenvectors of two matrices associated with the set of pages.

Let M denote the matrix whose (i,j) entry gives the number of pages that point to both page i and page j. Let N denote the matrix whose (i,j) entry gives the number of pages that are pointed to by both page i and page j.

Then the above iterative procedures are in fact an implementation of the power iteration method, given in G. Golub, C. F. Van Loan, "Matrix Computations", Johns Hopkins University Press, 1989, p. 351, for computing the principal eigenvectors of the matrices M and N.

In particular, the authority vector A is the principal eigenvector of M, and the hub vector H is the principal eigenvector of N.

The additional vectors A_i and H_i correspond to non-principal eigenvectors of M and N respectively. The use of such eigenvectors for clustering is known as spectral partitioning, and has been studied as a graph algorithm. See, for instance, D. Spielman, S. Teng, "Spectral partitioning works: Planar graphs and finite-element meshes," Proceedings of the 37th IEEE Symposium on Foundations of Computer Science, 1996.

The entries of the matrices M and N correspond to co-citation and bibliographic coupling, which have been studied in the bibliometric literature.

Since the algorithm works on an arbitrary set of linked pages, it is worth noting that it can be run in a query-independent fashion. In particular, given a set of pages, the identification of hubs and authorities among them gives a method for automatically determining the topic that best "fits" the set of pages.

SUMMARY

Using the foregoing specification, the invention may be implemented using standard programming and/or engineering techniques using computer programming software, firmware, hardware or any combination or subcombination thereof. Any such resulting program(s), having computer readable program code means, may be embodied or provided within one or more computer readable or usable media such as fixed (hard) drives, disk, diskettes, optical disks, magnetic tape, semiconductor memories such as read-only memory (ROM), etc., or any transmitting/receiving medium such as the Internet or other communication network or link, thereby making a computer program product, i.e., an article of manufacture, according to the invention. The article of manufacture containing the computer programming code may be made and/or used by executing the code directly from one medium, by copying the code from one medium to another medium, or by transmitting the code over a network.

An apparatus for making, using, or selling the invention may be one or more processing systems including, but not limited to, a central processing unit (CPU), memory, storage devices, communication links, communication devices, servers, I/O devices, or any subcomponents or individual parts of one or more processing systems, including software, firmware, hardware or any combination or subcombination thereof, which embody the invention as set forth in the claims.

User input may be received from the keyboard, mouse, pen, voice, touch screen, or any other means by which a human can input data to a computer, including through other programs such as application programs.

One skilled in the art of computer science will easily be able to combine the software created as described with appropriate general purpose or special purpose computer hardware to create a computer system and/or computer subcomponents embodying the invention and to create a computer system and/or computer subcomponents for carrying out the method of the invention. While the preferred embodiment of the present invention has been illustrated in detail, it should be apparent that modifications and adaptations to that embodiment may occur to one skilled in the art without departing from the spirit or scope of the present invention as set forth in the following claims.

While the preferred embodiments of the present invention have been illustrated in detail, it should be apparent that modifications and adaptations to those embodiments may occur to one skilled in the art without departing from the scope of the present invention as set forth in the following claims.

What is claimed is:

1. A computer program product, for use with a computer system, for directing the computer system to execute a search of information resources, the resources having content-based links between each other, to identify a desired subset of the information resources which satisfy a desired criterion, the computer program product comprising:

a computer-readable medium;

means, provided on the recording medium, for directing the computer system to identify an initial set of information resources;

means, provided on the recording medium, for directing the computer system to define initial authoritativeness information for the initial set;

means, provided on the recording medium, for directing the computer system to use the initial authoritativeness information as input authoritativeness information, to execute the steps of:

(i) producing first authoritativeness information about a set of information resources pointed to by links in resources of the input set, and

(ii) producing second authoritativeness information about a set of information resources having links that point to resources of the input set; and

means, provided on the recording medium, for directing the computer system to produce a final set of information resources based on the first and second authoritativeness information.

2. A computer program product as recited in claim 1, wherein the information resources include World Wide Web pages, and the content-based links include hyperlinks.

3. A computer program product as recited in claim 1, wherein the means for directing to identify an initial set of information resources includes means, provided on the recording medium, for directing the computer system to obtain, as an input, an information resource containing subject matter of interest.

4. A computer program product as recited in claim 3, wherein the means for directing to identify an initial set of information resources includes means, provided on the recording medium, for directing the computer system to identify a further set of information resources linked to the input information resource.

5. A computer program product as recited in claim 1, wherein:

the means for directing to execute the steps of producing first and second authoritativeness information is operative in a series of iterations;

the initial authoritativeness information is used as input authoritativeness information for a first iteration; and the produced first and second authoritativeness information is a result of the iteration, the first and second authoritativeness information produced in a given iteration to be used as the input authoritativeness information for the next iteration.

6. A computer program product as recited in claim 1 further comprising means, provided on the recording medium, for directing the computer system to execute the steps of producing first authoritativeness information and producing second authoritativeness information in a series of iterations until a predetermined condition is met.

7. A computer program product as recited in claim 6, wherein the predetermined condition includes the execution of a specified number of iterations.

8. A computer program product as recited in claim 6, wherein the predetermined condition includes a steady state in which further iterations result in substantially the same results.

9. A computer program product as recited in claim 6, wherein the means for directing to identify an initial set of information resources includes means, provided on the recording medium, for directing the computer system to execute a keyword-based query search, results of the search including information resources to be included in the initial set.

10. A computer program product as recited in claim 9, wherein the means for directing to identify an initial set of information resources further includes means, provided on the recording medium, for directing the computer system to identify information resources linked to or from the information resources which are the results of the search, the former information resources also to be included in the initial set.

11. A computer program product as recited in claim 10, wherein the means for directing to define initial authoritativeness information includes means, provided on the recording medium, for directing the computer system to select an initial numerical authoritativeness value for each of the information resources of the initial set.

12. A computer program product as recited in claim 11, wherein the means for directing to define initial authoritativeness information further includes means, provided on the recording medium, for directing the computer system to define an authority value and a hub value for each of the information resources of the initial set.

13. A computer program product as recited in claim 12, wherein the defined authority values and hub values are processed as vectors, each vector containing a respective term corresponding with each respective one of the information resources of the initial set, and having stored therein the value defined for that respective one of the information resources of the initial set.

14. A computer program product as recited in claim 12, wherein:

an initial hub value is defined as 1 if the information resource was found by the keyword-based query search, and 0 if the information resource is linked to or from the information resources which are the results of the search; and

an initial authority value is defined as 0 for all information resources.

15. A computer program product as recited in claim 12, wherein, for each iteration:

the hub value for an information resource is updated as the sum of the authority values for authority information resources which point to the hub information resource; and

the authority value for an information resource is updated as the sum of the hub values for hub information resources which are pointed to by the information resource.

16. A computer program product as recited in claim 15, wherein each iteration further includes normalizing the hub and authority values for the information resources.

17. A computer program product as recited in claim 1, wherein the means for directing to produce a final set of information resources includes means, provided on the recording medium, for directing the computer system to select information resources from the set based on their hub and authority values.

18. A computer program product as recited in claim 17, wherein the means for directing to select includes means, provided on the recording medium, for directing the computer system to select information resources whose hub values or authority values have greatest magnitudes.

19. A computer program product as recited in claim 17, wherein the means for directing to select includes means, provided on the recording medium, for directing the computer system to select a plurality of successive communities, selecting each successive community including selecting information resources whose hub values or authority values have greatest magnitudes of those information resources not already selected for a prior community.

20. A method for executing a search of information resources, the resources having content-based links between each other, to identify a desired subset of the information resources which satisfy a desired criterion, the method comprising the steps of:

identifying an initial set of information resources;

defining initial authoritativeness information for the initial set;

using the initial authoritativeness information as input authoritativeness information, executing the steps of:

(i) producing first authoritativeness information about a set of information resources pointed to by links in resources of the input set, and

(ii) producing second authoritativeness information about a set of information resources having links that point to resources of the input set; and

producing a final set of information resources based on the first and second authoritativeness information.

21. A method as recited in claim 20, wherein the information resources include World Wide Web pages, and the content-based links include hyperlinks.

22. A method as recited in claim 20, wherein the step of identifying an initial set of information resources includes obtaining, as an input, an information resource containing subject matter of interest.

23. A method as recited in claim 22, wherein the step of identifying an initial set of information resources includes identifying a further set of information resources linked to the input information resource.

24. A method as recited in claim 20, wherein:

the step of executing the steps of producing first and second authoritativeness information is executed in a series of iterations;

the initial authoritativeness information is used as input authoritativeness information for a first iteration; and

the produced first and second authoritativeness information is a result of the iteration, the first and second authoritativeness information produced in a given iteration to be used as the input authoritativeness information for the next iteration.

25. A method as recited in claim 20, wherein the steps of producing first authoritativeness information and producing second authoritativeness information are executed in a series of iterations until a predetermined condition is met.

26. A method as recited in claim 25, wherein the predetermined condition includes the execution of a specified number of iterations.

27. A method as recited in claim 25, wherein the predetermined condition includes a steady state in which further iterations result in substantially the same results.

28. A method as recited in claim 25, wherein the step of identifying an initial set of information resources includes executing a keyword-based query search, results of the search including information resources to be included in the initial set.

29. A method as recited in claim 28, wherein the step of identifying an initial set of information resources further includes identifying information resources linked to or from the information resources which are the results of the search, the former information resources also to be included in the initial set.

30. A method as recited in claim 29, wherein the step of defining initial authoritativeness information includes selecting an initial numerical authoritativeness value for each of the information resources of the initial set.

31. A method as recited in claim 30, wherein the step of defining initial authoritativeness information further includes defining an authority value and a hub value for each of the information resources of the initial set.

32. A method as recited in claim 31, wherein the defined authority values and hub values are processed as vectors, each vector containing a respective term corresponding with each respective one of the information resources of the initial set, and having stored therein the value defined for that respective one of the information resources of the initial set.

33. A method as recited in claim 31, wherein:

an initial hub value is defined as 1 if the information resource was found by the keyword-based query search, and 0 if the information resource is linked to or from the information resources which are the results of the search; and

an initial authority value is defined as 0 for all information resources.

34. A method as recited in claim 31, wherein, for each iteration:

the hub value for an information resource is updated as the sum of the authority values for authority information resources which point to the hub information resource; and

the authority value for an information resource is updated as the sum of the hub values for hub information resources which are pointed to by the information resource.

35. A method as recited in claim 34, wherein each iteration further includes normalizing the hub and authority values for the information resources.

36. A method as recited in claim 20, wherein:

each information resource is associated with an authority value and a hub value; and

the step of producing a final set of information resources includes selecting information resources from the set based on the hub and authority values.

37. A method as recited in claim 36, wherein the step of selecting includes selecting information resources whose hub values or authority values have greatest magnitudes.

38. A method as recited in claim 36, wherein the step of selecting includes selecting a plurality of successive communities, selecting each successive community including selecting information resources whose hub values or authority values have greatest magnitudes of those information resources not already selected for a prior community.

39. A system for executing a search of information resources, the resources having content-based links between each other, to identify a desired subset of the information resources which satisfy a desired criterion, the system comprising:

means for identifying an initial set of information resources;

means for defining initial authoritativeness information for the initial set;

means for using the initial authoritativeness information as input authoritativeness information, to execute the steps of:

(i) producing first authoritativeness information about a set of information resources pointed to by links in resources of the input set, and

(ii) producing second authoritativeness information about a set of information resources having links that point to resources of the input set; and

means for producing a final set of information resources based on the first and second authoritativeness information.

40. A system as recited in claim 39, wherein the information resources include World Wide Web pages, and the content-based links include hyperlinks.

41. A system as recited in claim 39, wherein the means for identifying an initial set of information resources includes means for obtaining, as an input, an information resource containing subject matter of interest.

42. A system as recited in claim 41, wherein the means for identifying an initial set of information resources includes

means for identifying a further set of information resources linked to the input information resource.

43. A system as recited in claim 39, wherein:

the means for executing the steps of producing first and second authoritativeness information is operative in a series of iterations;

the initial authoritativeness information is used as input authoritativeness information for a first iteration; and

the produced first and second authoritativeness information is a result of the iteration, the first and second authoritativeness information produced in a given iteration to be used as the input authoritativeness information for the next iteration.

44. A system as recited in claim 39 further comprising means for executing the steps of producing first authoritativeness information and producing second authoritativeness information in a series of iterations until a predetermined condition is met.

45. A system as recited in claim 44, wherein the predetermined condition includes the execution of a specified number of iterations.

46. A system as recited in claim 44, wherein the predetermined condition includes a steady state in which further iterations result in substantially the same results.

47. A system as recited in claim 44, wherein the means for identifying an initial set of information resources includes means for executing a keyword-based query search, results of the search including information resources to be included in the initial set.

48. A system as recited in claim 47, wherein the means for identifying an initial set of information resources further includes means for identifying information resources linked to or from the information resources which are the results of the search, the former information resources also to be included in the initial set.

49. A system as recited in claim 48, wherein the means for defining initial authoritativeness information includes means for selecting an initial numerical authoritativeness value for each of the information resources of the initial set.

50. A system as recited in claim 49, wherein the means for defining initial authoritativeness information further includes means for defining an authority value and a hub value for each of the information resources of the initial set.

51. A system as recited in claim 50, wherein the defined authority values and hub values are processed as vectors, each vector containing a respective term corresponding with each respective one of the information resources of the initial set, and having stored therein the value defined for that respective one of the information resources of the initial set.

52. A system as recited in claim 50, wherein:

an initial hub value is defined as 1 if the information resource was found by the keyword-based query search, and 0 if the information resource is linked to or from the information resources which are the results of the search; and

an initial authority value is defined as 0 for all information resources.

53. A system as recited in claim 50, wherein, for each iteration:

the hub value for an information resource is updated as the sum of the authority values for authority information resources which point to the hub information resource; and

the authority value for an information resource is updated as the sum of the hub values for hub information resources which are pointed to by the information resource.

54. A system as recited in claim 53, wherein each iteration further includes normalizing the hub and authority values for the information resources.

55. A system as recited in claim 39, wherein the means for producing a final set of information resources includes means for selecting information resources from the set based on their hub and authority values.

56. A system as recited in claim 55, wherein the means for selecting includes means for selecting information resources whose hub values or authority values have greatest magnitudes.

57. A system as recited in claim 55, wherein the means for selecting includes means for selecting a plurality of successive communities, selecting each successive community including selecting information resources whose hub values or authority values have greatest magnitudes of those information resources not already selected for a prior community.

* * * * *